

# VISUAL DATA FUSION FOR DECISION SUPPORT APPLICATIONS OF NUMERICAL WEATHER PREDICTION

Lloyd A. Treinish\*

IBM Thomas J. Watson Research Center, Yorktown Heights, NY

## 1. INTRODUCTION

Visualization is a method of computing by which the enormous bandwidth and processing power of the human visual (eye-brain) system becomes an integral part of extracting knowledge from complex data. In that regard, our previous work has discussed methods of appropriate mapping of user goals to the design of pictorial content by considering both the underlying data characteristics and the (human) perception of the visualization (Treinish, 1999a). The introduction of new applications further challenges these ideas.

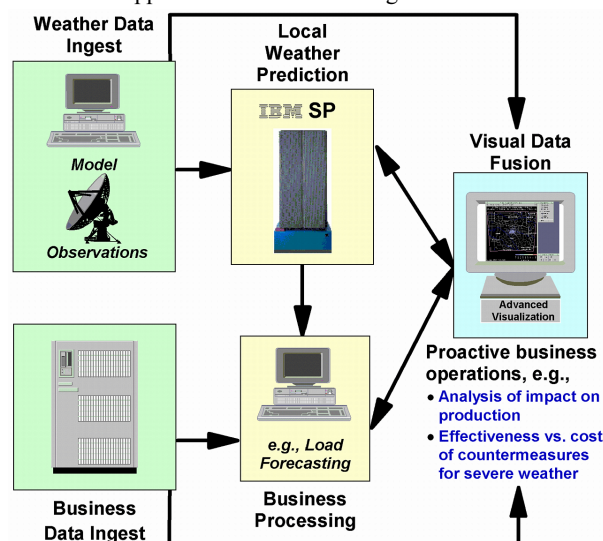


Figure 1. Visual data fusion for weather applications.

In particular, we have extended our earlier work (e.g., Snook et al, 1998) for situations where high-resolution models can be utilized in variety of weather-sensitive decision-making efforts such as emergency planning, energy production, airline operations, risk assessment, agricultural activities, commodity trading, etc. For each of these applications, information is assessed and decisions are made based upon a variety of static and dynamic data sets, a subset of which are weather-related. The utilization of these data and the complexity of the decision-making process changes when high-resolution predictive data are incorporated. These applications imply the coupling of weather simulations with other models, analyses and data. Visualization is a critical component to such integration. To enable effective assessment and appropriate decisions, focused visualizations must be designed to integrate these distinct data sources, yet still be driven by user goals. Resultant visualizations which represent a fusion of weather and non-weather data may not even illustrate forecasts of weather phenomena directly. In these cases, the relevant information is in the impact of weather via derived properties, which are influenced by weather, not weather variables produced by a simulation.

\* Corresponding author address: Lloyd A. Treinish, IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598, lloyd@us.ibm.com, <http://www.research.ibm.com/people/l/lloyd>

The problem is illustrated schematically in Figure 1. Two traditional data generators are shown on the top and the bottom (weather and non-weather, respectively). Although visualization is applicable to both, typically this is mutually independent. We propose an approach of visual data fusion to address the visualization design problem in such applications.

## 2. DATA FUSION

Data fusion is simply the integration of multiple data sets. This notion is derived from the fact that understanding of phenomena from a scientific basis, creating an engineering design, or assessment for sound decision making requires the utilization of data from many distinct sources. Traditionally such tasks have utilized a single data set, but as a result is often incomplete for larger-scale problems that are becoming more prevalent today. In parallel with the growth in problem complexity are additional factors that make the need for data fusion more practical and thus, more pervasive. The relative availability of relevant data enables a comparison study for a data generator as much as it does an independent analysis. Secondly, data generators have become more capable and accessible. Digital data acquisition is easier and cheaper. Computational simulations are gaining fidelity and detail while becoming more practical to compute. From verification of computational and experimental models to steering simulations with real-world observations, bringing data from multiple sources together is much more powerful than using each source separately. Visualization is critical to this integration, without which the beneficiaries of such data would be overwhelmed by volume or complexity (Usselton et al, 1998).

Data from multiple sources require care in their presentation so that artifacts due to the visualization process are not introduced by data fusion and erroneously interpreted as features in the data. For example, the data may not be uniformly available for the spatial domains being examined. Each of the data sets to be "fused" are generally not geographically co-registered and are defined on differing geometric structures. Further, the coordinate system for visualization and interaction may need to differ from those native to the data sets of interest.

These issues have been considered by others in a variety of applications including earth science, physics, astronomy and medical imaging (Usselton et al, 1998). In the majority of these cases, the user goals focused on analysis or verification as opposed to data assessment as illustrated in efforts to compare computational fluid dynamics results with experimental data from wind tunnels (Keely and Usselton, 1998). More recent work has considered decision support (Bisantz et al, 1999) but from a human factors perspective.

## 3. APPROACH

To enable visual data fusion, a perspective of data management must be adopted by introducing a uniform data model that is matched to the structure of the data as well how such data are used. This implies a generalized mechanism to classify and access data as well as efficiently map data to operations. The implementation of such a data model effectively decouples the management of and access to the data from the

actual application. This encapsulates the variety of sampling and representations for diverse data and provides uniform access. It is then a prerequisite to building applications that utilize the data sets to be integrated (Treinish, 1999b). One consequence of such a data (model)-centric approach is that the same operation(s) can be applied to data sets that need to be visually fused or correlated (i.e., displayed and interacted together) without introducing superfluous interpolation or resampling to a common mesh. The latter process implies a modification to the data, whose impact could be hidden in subsequent visualization. Further, if a specific visualization task requires a cartographic projection, then these data sets can be independently warped by the prerequisite transformation. Any geometric distortion that is introduced is due only to the actual projection since the data and topology remain invariant through such a transformation. It is also independent of the choice of realization or rendering technique or cartographic projection, and hence, provides a framework for experimenting with different visualization strategies. As a result, the fidelity of the original data sets is preserved in a coordinate system suitable for dynamic interaction. It implies that correlative visualization for visual fusion can be approached from four perspectives. In all cases, the specific choices are dictated by the goal of the visualization task(s) as defined by the individuals or applications utilizing the data.

1. *Image Level*. The capability to look at multiple sets of data in exactly the same fashion (i.e., visual comparison within a common framework). This can be achieved with multiple visualizations in adjacent windows or mosaiced together for qualitative comparison. These visualizations are usually static, but might be accompanied by synchronized animation sequences or geometric transformations in which the representations are linked. Outside of the latter, interaction is typically indirect.
2. *Common View*. The capability to utilize a variety of visualization strategies within a chosen coordinate system dictated by one of the data sets or independently by user task. This represents a visual fusion which can support both direct and indirect interaction, including numerical querying. All of the relevant data are registered within this common viewing framework. Qualitative comparisons are clearly supported, but direct quantitative comparisons are defined by interaction.
3. *Data Level*. The capability to numerically compare distinct data sets using either of the two previous approaches for visualization. This does require the transformation (e.g., interpolation) of one or more data sets to a common basis (mesh, coordinate system, etc.) from which derived quantities can be calculated (e.g., point-wise opera-

tions). The visualizations may involve the original data and/or the derived data. From the discussion earlier this can violate the principle of preserving fidelity at the cost of supporting numerical comparisons.

4. *Multiple Views*. The capability to numerically and visually compare multiple data sets, particularly when some of the data sets do not have a common basis for visual fusion. In this case, the utilization of a variety of different strategies is required, some of which must be in separate instead of a common framework. Interaction may be complex because separate metaphors for direct interaction are required for each framework, although common methods for indirect interaction are feasible. Unlike case 1, quantitative access is supported such that linked displays would indicate related numerical values or "regions" of commonality that are queried.

## 4. APPLICATIONS

The aforementioned approach to data fusion is applied to problems that relate to economic and societal impacts of weather. In some cases, additional sources of weather information, both historical and predictive may have further benefit, when integrated into the decision-making process.

### 4.1 Emergency Planning

Weather-related catastrophes have led to over \$48B in property insurance claims from 1989 to 1993 in the US. In North Carolina alone, ten major hurricanes from 1983 to 1996 resulted in about \$50B worth of damage, almost \$30B of which led to losses by insurance companies (Kunkel et al, 1999). Hence, disaster planning or hedging for underwriting risk-related insurance can benefit from improved weather predictions. In both cases, the impact of weather is relevant in visualization but not the weather data directly. Although geo-referenced visualizations are required, the illustration of time-dependent factors related to property loss due to severe weather are needed, not merely a visualization of predicted wind velocity, for example. Usually, an *Image Level* (case 1) approach is applied as shown in Figure 2. Each image contains a simple two-dimensional map of a set of glyphs colored by a different parameters. The glyphs are located at the centroid of the area associated with zip codes.

However, the glyph locations are only marked on the map when a set of conditions on house value, population and estimated damage due to wind are met. Therefore, a *Common View* (case 2) approach is more efficient by leveraging user interaction as illustrated in Figure 3. The user is free to interactively set the conditions and animate in time corresponding to the weather simulation in hourly steps. This enables the determination of areas of greatest impact due to severe

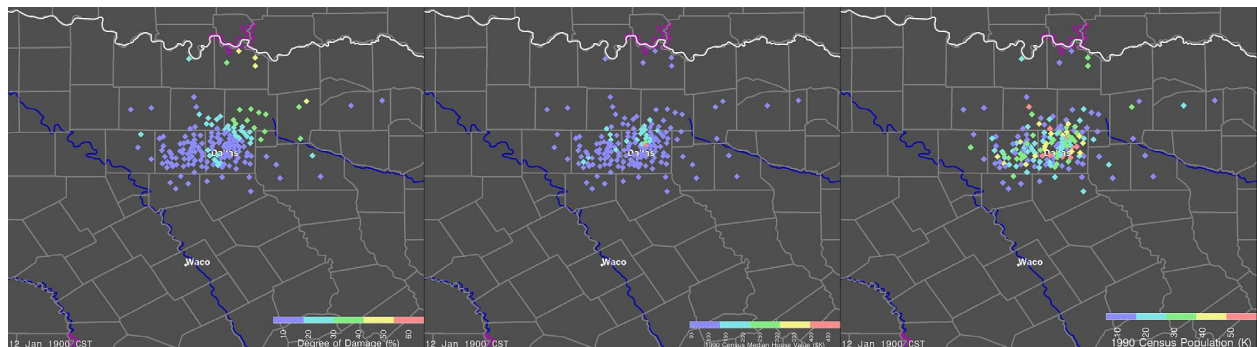
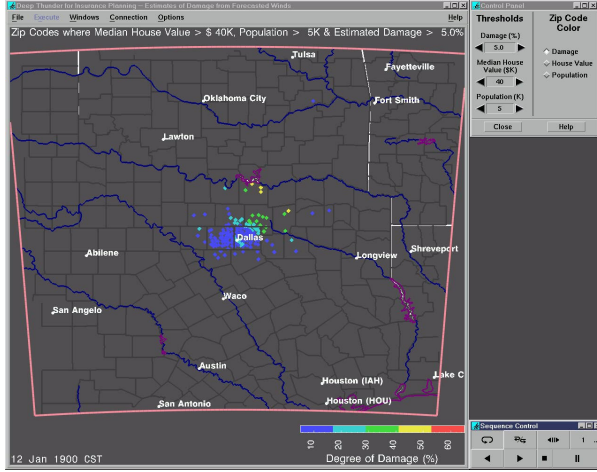


Figure 2. *Image Level* data fusion of a weather forecast with demographic data over an 800 x 800 km domain at 8 km resolution centered over Dallas. Colored glyphs at zip code locations illustrate a subset of demographic and derived data.

weather. Essentially, it represents a simple method to specify a query against various data sets, which are then used to constrain a visual integration for display and interaction. This approach becomes *Data Level* (case 3) because the forecast data are interpolated to zip code locations in order to support the query constraints. These thresholds can also be augmented to include other relevant demographic, customer or property data. The demographic data shown are derived from available census information (<http://tiger.census.gov>).



**Figure 3. Common View** data fusion showing the relationship between demographic data and a weather forecast in a screen capture of an interactive session. This also is a *Data View* data fusion because the locations of estimated damage are calculated from the weather model data.

In this example, the conditions for display are enhanced to include a simple computational model. The level of wind-induced damage is based upon analysis of effects on typical residential buildings from severe weather (Unanwa et al, 2000). This approach to data fusion may be useful for planning purposes by an insurance company or deployment of repair crews by a utility or local highway department.

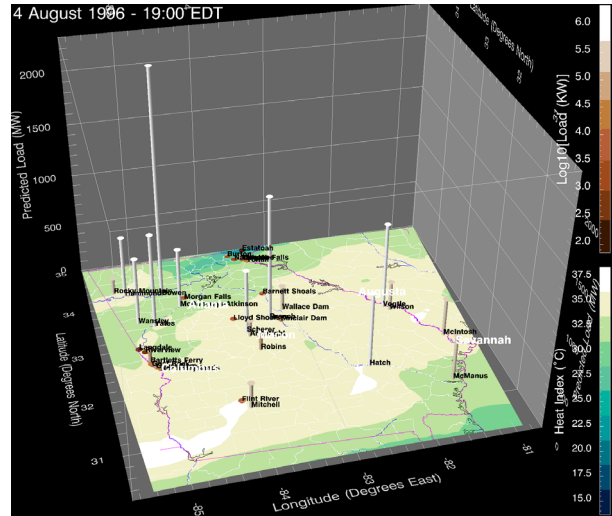
#### 4.2 Energy Demand Forecasting

Another application of a predictive weather model is to forecast load on a power-generation facility or transmission lines for efficient running of the facility or for power trading. In both cases, meteorological information is an important input as weather is a primary driver for electricity demand. It has been estimated that the annual cost of under or over predicting electricity demand due to poor temperature forecasts is several hundred million dollars in the US alone. Erroneous weather data associated with startup-shutdown of generation units can be worth \$500K per day during peak load periods or conservatively \$8M annually to a regional power authority. In addition, improved severe storm predictions to reduce outage time can save a few hundred thousand dollars a year for a typical utility (Keener, 1997). Decisions in this industry are driven by diverse non-weather data and processes including load forecasting and econometric models, customer demographics, geography of power facilities, etc. that are not well integrated. The weather information currently used is relatively coarse leading to poor and costly decisions. Typically, hourly forecast surface temperature and dew point values averaged over a large geographic region are used. Alternatively, more accurate data at greater frequency which are distinct for different loads by geographic location and altitude can be applied coupled with other factors that influence load (e.g.,

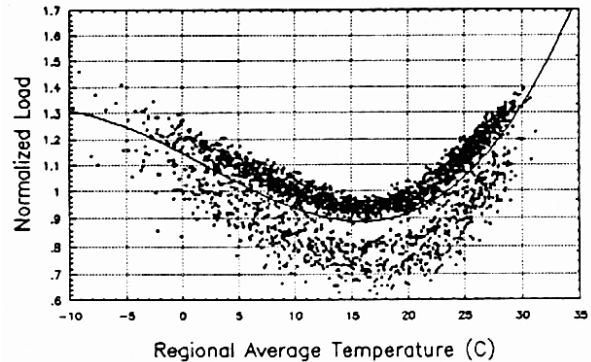
storm and cloud predictions). Since there is a relationship between accuracy in load prediction vs. economic efficiency (i.e., an under prediction implies having to buy power at a premium and over prediction means resources are wasted), coupling of weather forecasts with econometric models is also feasible.

These ideas are illustrated in Figure 4 using *Data Level* fusion. It shows a map of Georgia with forecasted heat indices at 8 km resolution. Major cities and locations of the generators owned and operated by Georgia Power, the local electric utility, are shown by name. Each power plant location is also marked with a pin, whose height and color indicate a predicted electricity demand. A dual encoding is used because the capacities of the power plants range over five orders of magnitude. Hence, height is a linear mapping while color bands are scaled logarithmically. The load is computed interactively as a function of temperature, humidity and time of day from a simple model. The temperature dependence is based upon a polynomial approximation of the relationship between historical data of power demand and weather observations, shown in Figure 5 (Robinson, 1997). Regression on the data from summer weekdays in the southeastern United States after outliers are removed yields,

$$W = 1.146 - 0.0225T - 0.000240T^2 + 0.0000397T^3 \quad (1)$$



**Figure 4. Data Level** data fusion of a weather forecast at 8 km resolution centered over Atlanta with a prediction of electricity demand at power plants operated by Georgia Power. The demand is calculated from a model whose input is derived from the numerical weather prediction.



**Figure 5. Weather-dependent energy load,  $W(T)$ .**



The temporal variation is based upon a spline fit of hourly electricity requirements for mid-week days in urban and sub-urban tropical environments, which is consistent with other results in the literature (Chang and Li, 1998). That component is shown in Figure 6, which is then normalized for this application.

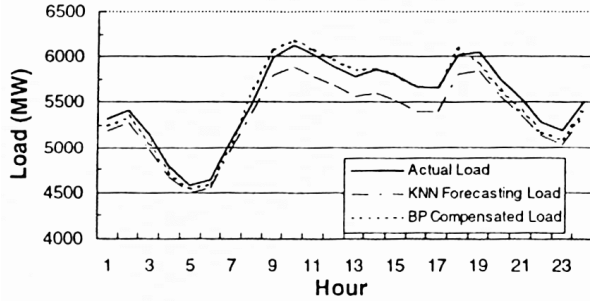


Figure 6. Diurnal, mid-week energy load,  $N(t)$ .

The temperature and temporal components are combined for a total estimated load,  $L$ , such that

$$L = C[(0.2768809N(t) + 0.7231191)(W(T_{HI})/2.9175)] \quad (2)$$

The function is scaled by the rated power plant capacity,  $C$ , using published data (<http://www.georgiapower.com/newsroom/plants.asp>). Heat index,  $T_{HI}$ , is employed as a more accurate measure of demand than simply temperature. It is an apparent temperature derived from both temperature and humidity as an indicator of personal comfort during the summer (Rothfusz, 1990). Therefore, it is directly related to air conditioning usage and thus, electricity demand. The weather model results are interpolated at each time step to the location of each of the power plants. An example for the specific 24-hour period of the forecast is shown in Figure 7 for the largest power plant operated by Georgia Power (Bowen).

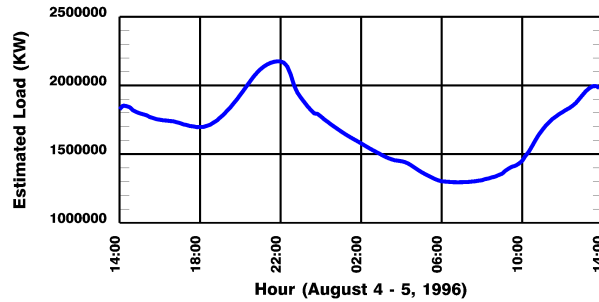


Figure 7. Predicted power demand at a specific generator site derived from a weather-model-driven load forecast.

All of these capabilities are illustrated in Figure 8, which is a screen capture of a prototype of an interactive application for detailed load forecasting. The user has the ability to select the type of power plant (fossil, hydroelectric and/or nuclear), what data to show on the map (e.g., weather, geographic or other customer/demographic) and to query individual power plants (i.e., by visual selection). The results of the query include the predicted load at each time step (as fine as every 10 minutes) as well as a plot of predicted load over 24 hours with weather data at that location. The interactive application is then a *Multiple View* (case 4) fusion.

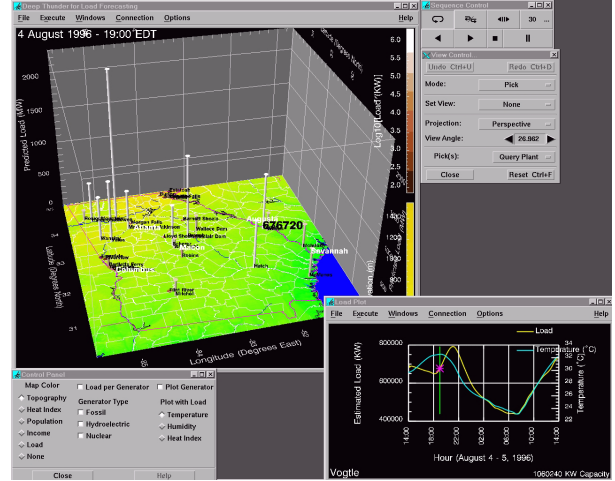


Figure 8. *Data Level* (3d window) and *Multiple View* (with 2d plot) data fusion for weather-model-driven energy load forecasting in a screen capture of an interactive session. The multiple views are linked in time sequence and by interactive selection in the 3d window.

The visual fusion techniques of Figures 3 and 4 are combined in Figure 9, which shows the load forecast at the power plants that use fossil fuels with a population map. The population data are shown as colored contours on a logarithmic scale to segment urbanized areas (red) and their location with respect to the power generators under the heaviest demand. Although these data are derived from static census sources, the same techniques would apply to similar but proprietary customer data owned by a regional electric utility.

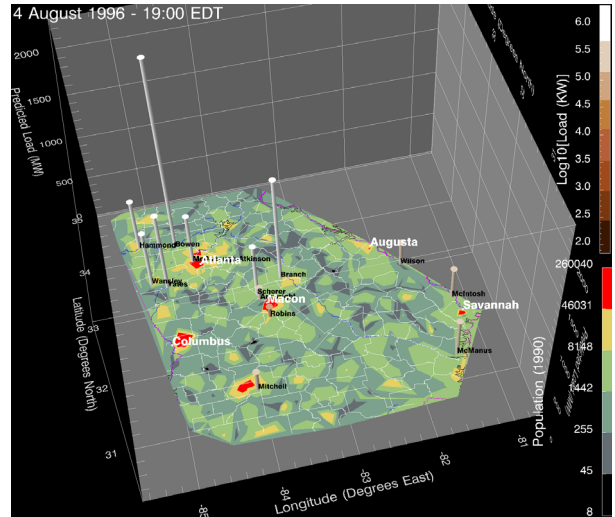


Figure 9. *Data Level* and *Common View* data fusion to illustrate the correlation of a load forecast with demographic data. The proximity of fossil fuel power plants with high predicted load to major population centers (red) is easily seen.

## 5. IMPLEMENTATION

The applications shown in Figures 3 and 8 present a user interface based upon XWindow/Motif for indirect interaction and OpenGL for direct three-dimensional interaction in cartographic coordinates native to the weather simulation. They have been implemented with Data Explorer (DX) (Abram and

Treinish, 1995). DX is a portable, open source, general-purpose software package for visualization and analysis (<http://www.research.ibm.com/dx> and <http://www.opendx.org>). A generic toolkit was used to avoid having to implement a graphics and computational infrastructure. Unlike traditional meteorological graphics or geographic information systems, DX is parallelized for multiprocessor workstations and can utilize three-dimensional graphics accelerators. DX is built upon an unified data model that enables these applications to operate directly on the native gridded weather data without transformation or compression.

## 6. CONCLUSIONS AND FUTURE WORK

The visualization of applications of high-resolution weather modelling have benefited from a focus on specialized interfaces and tools matched to user goals and underlying visualization tasks. Since the underlying toolkit is extensible tools can be reused between applications with similar user interface components. Although these applications and associated user goals are different, underlying data fusion requirements and visualization tasks are the same. Further, the need to employ a relatively simple user interface is desirable to reduce the effort for training of users in time-critical activities such as decision support. It also reduces the cost of development and maintenance, and enables more rapid iterative refinement with or adaptation to new users. Therefore, within any given application, incorporation of additional and more complex data sets can also be addressed. But the goal remains the same -- to develop simple interfaces and useful visual fusion.

The specific work discussed herein is on-going. To date, most of this work has been of a prototyping nature as both generic proof of concept as well as illustration of specific decision support problems. One aspect of continued efforts will be to incorporate more sophisticated models or processing as the consumer of weather forecast data. For example, the simple load forecasting model can be enhanced to include wind speed and sunshine duration effects and also, adjusted for more realistic temporal variation based upon day of week and season. To aid in the decision making applications, migration to a probabilistic representation will also be advantageous. In addition, it is believed that these ideas can be extended to other application areas such as agriculture, aviation, finance, etc.

The need for this type of data fusion occurs in other disciplines when an integrated view of the problem domain is adopted. Hence, the applicability of these methods to other problem areas will be investigated. This includes large-scale engineering design when computer-aided (mechanical) designs have to integrated with computational fluid dynamics, structural analysis, styling design, wind-tunnel measurements with physical models and actual testing with full-scale prototypes. Other applications may include multi-modal imaging such as in astronomy, earth sciences or medicine. Although challenging they have less data diversity because conceptually similar measurements are taken of the same physical space with varying sampling schemes that yield different informa-

tion that have to be registered into a common coordinate system.

## 7. REFERENCES

- Abram, G. and L. Treinish. *An Extended Data-Flow Architecture for Data Analysis and Visualization*. **Proceedings of the IEEE Visualization 1995 Conference**, October 1995, Atlanta, pp. 263-270.
- Bisantz, A. M., R. Finger, Y. Seong and J. Llinas. *Human Performance and Data Fusion Based Decision Aids*. **Proceedings of the FUSION '99 Conference**, July 1999, Sunnyvale, Volume 2, pp. 918-925.
- Chang, C. S. and M. Yi. *Real-Time Pricing Related Short-Term Load Forecasting*. **Proceedings of the Energy Management and Power Delivery 1998 Conference**, March 1998, Singapore, pp. 411-416.
- Keely, L. and S. Uselton. *Development of a Multi-Source Visualization Prototype*. **Proceedings of the IEEE Visualization 1998 Conference**, October 1998, Raleigh, pp. 411-414.
- Keener, R. N. *The Estimated Impact of Weather on Daily Electric Utility Operations*. **Proceedings of the Workshop on the Social and Economic Impacts of Weather**, April 1997, Boulder (<http://www.esig.ucar.edu/socasp/weather1/keener.html>)
- Kunkel, K. E., R. A. Pielke, Jr., S. A. Changnon. *Temporal Fluctuations and Climate Extremes that Cause Economic and Human Health Impacts: A Review*. **Bulletin of the American Meteorological Society**, **80**, n. 6, June 1999, pp. 1077-1098.
- Robinson, P. J. *Modeling Utility Load and Temperature Relationships for Use with Long-Lead Forecasts*. **Journal of Applied Meteorology**, **36**, n. 5, May 1997, pp. 591-598.
- Rothfusz, L. P. *The Heat Index Equation (or more than you ever wanted to know about heat index)*. **National Weather Service Southern Region Technical Attachment, SR/SSD 90-23**, Fort Worth, 1990.
- Snook, J. S., P. A. Stamus, J. Edwards, Z. Christidis, J. A. McGinley. *Local-Domain Mesoscale Analysis and Forecast Model Support for the 1996 Centennial Olympic Games*. **Weather and Forecasting**, **13**, n. 1, 1998, pp. 138-150.
- Treinish, L. *Task-Specific Visualization Design*. **IEEE Computer Graphics and Applications**, **19**, n. 5, September/October 1999, pp. 72-77.
- Treinish, L. *A Function-Based Data Model for Visualization*. **Proceedings of the IEEE Visualization 1999 Conference Late Breaking Hot Topics**, October 1999, San Francisco, pp. 73-76.
- Unanwa, C. O., J. R. McDonald, K. C. Mehta and D. A. Smith. *The Development of Wind Damage Bands for Buildings*. **Journal of Wind Engineering and Industrial Aerodynamics**, **84**, n. 1, January 2000, pp. 119-149.
- Uselton, S., J. Ahrens, W. Bethel, L. Treinish and A. State. *Multi-Source Data Analysis Challenges*. **Proceedings of the IEEE Visualization 1998 Conference**, October 1998, Raleigh, pp. 501-504.